DOCUMENT RESUME

ED 036 188 24 EM 007 774

AUTHOR UNKS, NANCY J.
TITLE A PLAN FOR EVALUATING THE IPI TESTING PROGRAM.
INSTITUTION RESEARCH FOR BETTER SCHOOLS, INC., PHILADELPHIA, PA.
SPONS AGENCY OFFICE OF EDUCATION (DHEW), WASHINGTON, D.C. BUREAU OF RESEARCH.
BUREAU NO BR-6-2867
PUB DATE 3 NOV 67
CONTRACT OEC-1-7-062867-3053
NOTE 9P.; PAPER PRESENTED AT PENNSYLVANIA EDUCATIONAL RESEARCH ASSOCIATION MEETING (UNIVERSITY OF PITTSBURGH, NOVEMBER 3, 1967)

EDRS PRICE EDRS PRICE MF-$0.25 HC-$0.55
DESCRIPTORS *INDIVIDUALIZED INSTRUCTION, *PROGRAM EVALUATION, TEST CONSTRUCTION, *TESTING, TESTING PROGRAMS, TEST INTERPRETATION, TEST RELIABILITY, TEST VALIDITY
IDENTIFIERS INDIVIDUALLY PRESCRIBED INSTRUCTION, IPI

ABSTRACT

THE TESTING SUB-PROGRAM IS DESIGNED TO PROVIDE THE DIAGNOSTIC INSTRUMENTS NECESSARY TO MEASURE PUPIL PROGRESS THROUGH THE INDIVIDUALLY PRESCRIBED INSTRUCTION (IPI) CURRICULA. ITS OBJECTIVES ARE TO PROVIDE INFORMATION ABOUT PUPILS WHICH TEACHERS CAN USE TO DIRECT EACH CHILD'S INDIVIDUAL LEARNING PROGRAM, TO PROVIDE THE MEASUREMENTS NECESSARY FOR THE RESEARCH AND EVALUATION OF THE IPI PROJECT AND ITS CURRICULA, AND TO PROVIDE INFORMATION ABOUT THE IPI ACHIEVEMENT TESTS THAT CAN BE USED TO IMPROVE THE TEST'S SERVICE TO TEACHERS AND PUPILS. THE EVALUATOR MUST BE ABLE TO DEFINE IN OPERATIONAL TERMS THE PURPOSE OF THE TEST, COMPARE THESE OBJECTIVES WITH THE PLAN FOR THE OPERATION OF THE TESTING PROGRAM, AND EVALUATE ITS ACTUAL OPERATION. THE ACTUAL ASSESSMENT OF THE TESTING PROGRAM'S OUTCOMES REQUIRES VALIDITY, RELIABILITY, AND ITEM ANALYSIS INFORMATION FOR ALL THE IPI INSTRUMENTS. AN OUTLINE OF THE EVALUATION PLAN FOR THE IPI TESTING PROGRAM IS PRESENTED IN DIAGRAM FORM. (JY)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED036188

A Plan for Evaluating the IPI Testing Program[1]

Nancy J. Unks

The Individually Prescribed Instruction (IPI) Project of the University of Pittsburgh's Learning Research and Development Center can be summarized, in brief, as a rather large scale educational innovation for individualizing instruction. Over the past three years, the problems of managing a growing program of this size and scope have led to the development of several sub-programs within the IPI system, each performing unique functions for the total program. This paper and the three that follow it are concerned with two of these sub-systems--the testing program and the computer management program. They also discuss how these sub-programs work together to perform the function of evaluating the total IPI Project.

First, let's look at the testing sub-program and what it does for IPI. Its primary function is to provide the diagnostic instruments necessary to measure pupil progress through the IPI curricula. In other words, the testing program produces the achievement tests that assess a pupil's mastery of specific behavioral objectives in math, reading, and science. In producing these instruments, the testing program performs several services for IPI which may be called its unique objectives. The testing program's objectives may be summarized as follows:

---

[1]Paper presented at the Pennsylvania Educational Research Association meeting, University of Pittsburgh, November 3, 1967.

EM001774

1. To provide information about pupils which teachers can use to direct each child's individual learning program (i.e., to provide diagnostic measurements meeting specified criteria for reliability, validity and item characteristics).
    a. Placement information
    b. Unit pre- and posttest information
    c. Information about pupil mastery of single skills
    d. Pupil rate of progress information
    e. Supplementary information about pupil aptitudes and abilities.
2. To provide the measurements necessary for the research and evaluation of the IPI Project and its curricula.
3. To provide information about the IPI achievement tests that can be used to improve the tests' service to teachers and pupils (i.e., to provide for testing program self-evaluation).
    a. Validity information
    b. Reliability information
    c. Item analysis data
    d. Reactions of IPI personnel to test formats, clarity, relevance, efficiency

It is quite necessary for the testing program to fulfill its first major objective or service in order for the total IPI project to reach its goals; teachers need the information the achievement tests provide in order to prescribe work for the pupils. The more complete the information available about each pupil's abilities and past achievement, the more accurately teachers will be able to prescribe instruction that fits the individual pupil's needs. This tailoring of instruction to each child's needs is the first goal of Individually Prescribed Instruction.

The second major function of the testing sub-system is a result of the emphasis that is placed on evaluation and research activities in the IPI Program. The testing program is responsible for obtaining those measurements of pupil progress which make possible the evaluation of the total IPI effort.

At this point, a rationale for evaluating the testing sub-program becomes apparent. The evaluation of the total project is dependent upon measurements made by the testing program, therefore, if these measurements are inadequate, little credence can be placed in an evaluation based on them. A researcher interested in the worth of the IPI Project as a whole first needs to know the worth of the measurements made upon it. In other words, evaluating the testing program is just as important as evaluating the total innovation of which it is part.

This rationale is the reason for the third major objective of the testing program. In fulfilling this goal, the sub-program provides information about itself which helps to answer the important question, "Can an evaluation of the total IPI Project based upon this testing program be considered sound?"

Once a need is established for evaluating the testing program, we begin to consider how to go about such a study. Many of the principles and procedures for evaluating educational programs in general can be applied, and, ideally, the testing program evaluation should be concurrent with and equal in magnitude to the total IPI evaluation.

The attached diagram presents an outline of the evaluation plan for the IPI testing sub-program. The large boxes show five major components of the testing porgram and its evaluation. The arrows between represent steps taken to assess how well these components fit or work togehter. Beneath the diagram are some of the questions a

researcher must ask to find out how well the program is operating and how well its parts function together. Each question is written under the step in the paradigm to which it applies. These questions are rather subjective, but the evaluator must try to answer them impartially and support his answers with objective evidence.

As a first step in evaluating the testing program, the researcher needs to know what the testing program is supposed to be accomplishing--He needs to have its objectives spelled out in quite unambiguous, operational terms. Having the objectives stated operationally, he then has some indication as to how their achievement can be assessed. For example, the first general objective listed on page 2 can be broken down into about seven operational goals which specify the desired functions of the IPI tests. The first of these would be, "To provide achievement tests specifically content-referenced to the behavioral objectives of the IPI curricula." To assess this goal, we can actually check whether such tests exist for each curriculum objective.

The second operational objective of the testing program would be, "To place pupils in proper work levels to begin study in IPI when they first enter an IPI school and at the beginning of each school year." Assessing the achievement of this objective is a little more complicated. We not only have to check whether placement tests do exist for all levels and units of work, but we also have to find out whether the tests place pupils in the proper work levels. An estimate of the validity of a placement test can help give an idea of how accurately it assesses pupil achievement. Concurrent validity might be obtained by administering, to a selected sample of students, both the placement test and the set of pretests covering the same units of work. Then we could compare the results of the placement tests with those of the pretests which supposedly

measure the same skills in greater detail. Another way to find out whether pupils have been placed properly is to examine their work patterns during the first two months of school and identify cases of misplacement. We have an indication that a pupil has been misplaced if he seems to have unusual difficulty with the work or if he goes through it with extreme ease. Spotting such cases of misplacement will be greatly facilitated by the computer management system that is being implemented for the IPI Project.

Other operational objectives of the testing program delineate the functions of unit pre- and posttests and curriculum embedded tests. The researcher, once he knows what these tests are supposed to accomplish, can then discern what statistical characteristics the tests should possess if they are fulfilling their desired functions. The third general objective on page 2 can be broken down into operational ones which elaborate the test characteristics the evaluator must look for in order to assess the worth of the IPI tests and hence of the IPI testing program. For example, placement instruments are designed to be general tests of many skills and, therefore, should have low internal consistency reliabilities and low inter-item correlations. Tests of single skills, on the other hand, are supposed to be quite homogeneous and should have high internal consistency and high inter-item correlations.

After obtaining operational objectives for the program, the evaluator next needs to examine some sort of specific plan for its operation. In this case, the plan for the testing program can be extracted from several papers describing the total IPI Project. The evaluator must review these sources critically, asking, "Does this plan show promise for accomplishing the testing program objectives, and does it outline the operation of the program in detail?"

The third component of the program to be studied is its actual operation. The evaluator must thoroughly define the testing program as it exists apart from the plan. He must elaborate on its personnel at the Learning Research and Development Center, its facilities at the Center and at Oakleaf School, and the achievement tests which are its primary products. The account of the existing testing program, together with the plan, consititutes the 'description of the innovation' that is essential to the evaluation of any educational program.

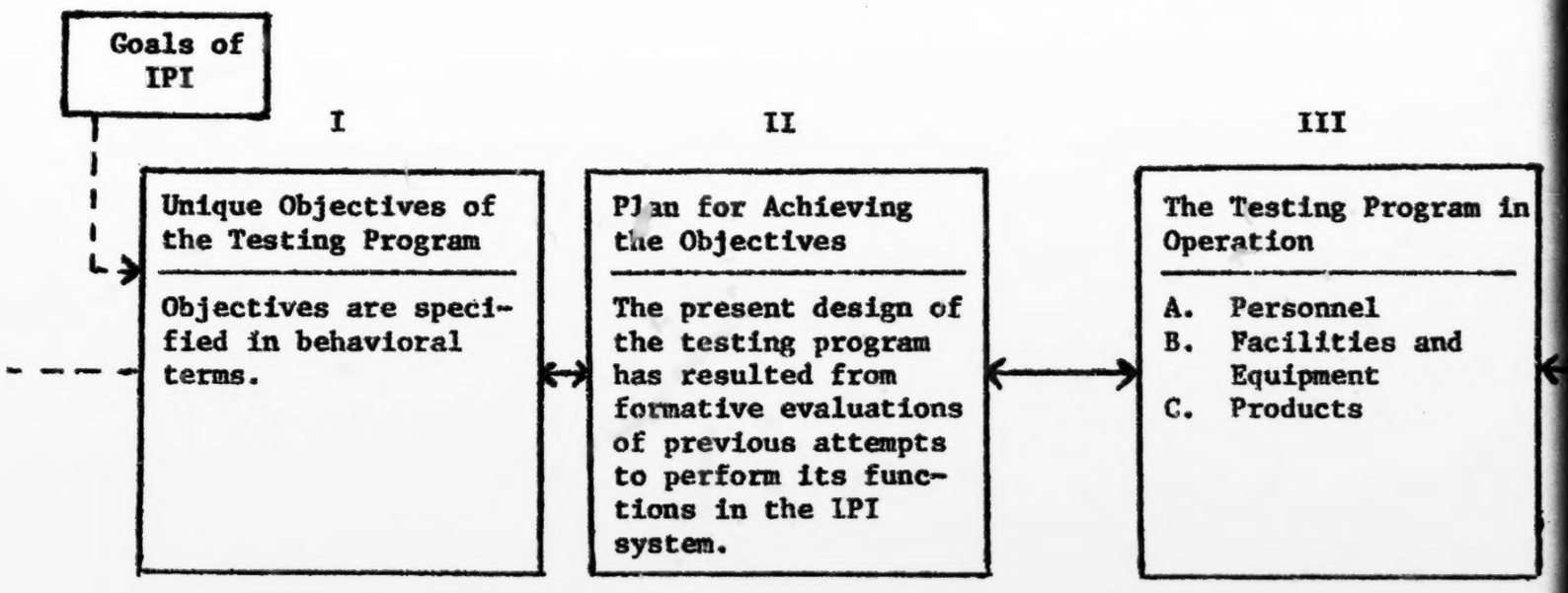
Finally, we get to what may be considered the heart of the evaluation--the actual assessment of the testing program's outcomes. We have already indicated how the achievement of its objectives will be measured; their assessment depends upon how the objectives themselves are stated. We can't just say that an objective of the testing program is to write "good" tests; we have to specify what constitutes a good test or, in other words, what that test should be able to do. A major portion of this plan for evaluating the IPI testing sub-system calls for finding out how well the tests fulfill their functions. This means the evaluator must obtain validity, reliability, and item analysis information for all of the IPI instruments and determine whether these meet acceptable standards as defined by the objectives. Obviously, collecting and organizing information like this for several hundred tests could be a monumental task. Most of this data collection can be handled automatically by the computer system that records the daily progress and test performance of every pupil at Oakleaf.

Along with the massive objective assessment of the testing program, the evaluator should also be concerned with the more subjective observation, description, and evaluation of unexpected or unplanned outcomes. He

must be alert to notice the effects of such things as breakdown in communications between test writers and lesson writers, delays in getting needed materials, and changes in the curriculum during the school year. All such observations should be reported regularly and explicitly.

At last the researcher, armed with a list of operational objectives, a definition of the testing program, and a mountain of data and observations on its outcomes, can begin to consider the possible implications of his evaluation. Does the assessment indicate that the objectives of the testing program have been achieved? Are unplanned results beneficial or unfavorable? Can implications drawn from the assessment be justified logically by the empirical evidence? Do the outcomes suggest needed changes in the objectives, plan and/or operation of the testing program? Does the testing program contribute to achievement of IPI goals? The entire rationale for evaluating the testing sub-program can be summarized in one question. Can an evaluation of the total IPI Project using measurements made by this testing program be considered meaningful? The answer to this question summarizes the completed evaluation.

Outline of the Evaluation Plan for the


Goals of IPI
I
II
III
Unique Objectives of the Testing Program
Objectives are specified in behavioral terms.
Plan for Achieving the Objectives
The present design of the testing program has resulted from formative evaluations of previous attempts to perform its functions in the IPI system.
The Testing Program in Operation
A. Personnel
B. Facilities and Equipment
C. Products

1. Are the testing program objectives relevant to the goals of IPI?

2. Are the objectives stated operationally?

3. Does the plan show promise for accomplishing the objectives?

4. Does the plan specify the operation in detail?

5. Does the program follow the plan?

6. Does the operation suggest needed changes in the plan?

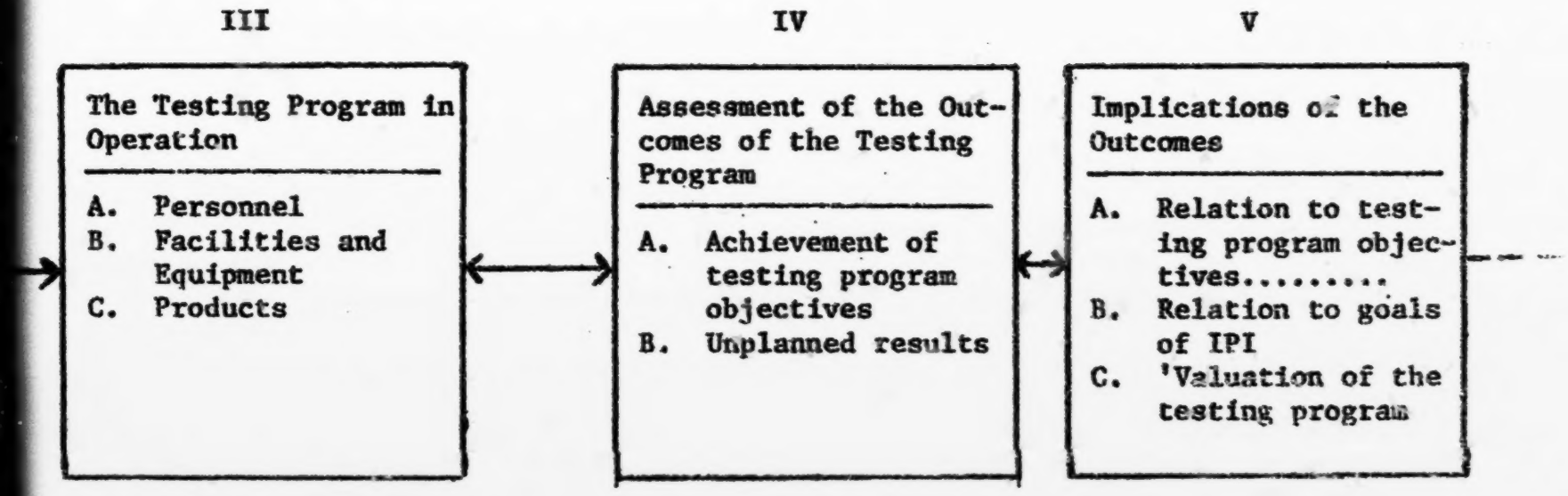
7. Does the program in operation follow logically from the testing program objectives?

8. Does the operation suggest needed changes in the objectives?

9. Does the testing program function efficiently in producing its products?

*Adapted from a general evaluation paradigm by C.M. Lindvall

the Evaluation Plan for the IPI Testing Program*



lan per- il?

ro- he

per-

s in

7. Does the program in operation follow logically from the testing program objectives?

8. Does the operation suggest needed changes in the objectives?

9. Does the testing program function efficiently in producing its products?

10. Is the assessment of outcomes objective?

11. Is the assessment of unplanned results comprehensive?

12. Does the assessment indicate that the objectives of the testing program have been achieved?

13. Are unplanned results beneficial or unfavorable?

14. Can implications drawn from the assessment be justified logically by the empirical evidence?

15. Do outcomes suggest needed changes in the objectives, plan and/or operation of the testing program?

16. Does the testing program contribute to achievement of IPI goals?

17. Can an evaluation of the total IPI Project using measurements made by this testing program be considered meaningful?